The Annals of Statistics
1986, Vol. 14, No. 3, 925–933

ASYMPTOTIC CONDITIONAL INFERENCE FOR THE OFFSPRING MEAN OF A SUPERCRITICAL GALTON–WATSON PROCESS

BY T. J. SWEETING

University of Surrey

Consider a supercritical Galton–Watson process (Z_n) with offspring distribution a member of the power series family, and having unknown mean θ. The conditional asymptotic normality of the suitably normalized maximum likelihood estimator of θ given the conditional information is established. The conditional information here is proportional to the total number of ancestors V_n, and it is also seen that this statistic is asymptotically ancillary for θ in a local sense. The proofs are via a detailed analysis of the joint characteristic function of (Z_n, V_n), and the derivation serves to highlight the difficulties involved in establishing such conditional results generally.

1. Introduction. This paper is concerned with asymptotic conditional inference for the offspring mean θ in a supercritical Galton–Watson branching process. The branching process with unknown mean is an instance of a *nonergodic* statistical model, where the appropriately normed sample Fisher information, $W_n(\theta)$ say, converges to a nondegenerate random variable W, rather than to a constant; see for example Basawa and Scott (1983). Under suitable regularity conditions it can be shown that, for such a model, $(X_n(\theta), W_n(\theta))$ converges in joint distribution to (Z, W) where $X_n(\theta)$ is the appropriate *randomly* normed maximum likelihood estimator (m.l.e) $\hat{\theta}_n$, and Z is a normal random variable independent of W [Sweeting (1980) and Basawa and Scott (1983)]. This result suggests that (a) some statistic V_n related to $W_n(\theta)$ might be regarded as *asymptotically ancillary* for θ, since the asymptotic distribution of $W_n(\theta)$ is continuous in θ, and hence effectively constant over the main range of variation of the distribution of $\hat{\theta}_n$, and (b) the *conditional* sampling distribution of $X_n(\theta)$ given V_n, whose use would be dictated by the conditionality principle, would still be asymptotically normal.

If this is the case, then approximate confidence intervals for θ based on this conditional distribution will coincide with approximate Bayesian h.p.d. intervals. Similar remarks have been made in Sweeting (1978, 1980) and amplified in Feigin and Reiser (1979). As noted in Sweeting (1982, 1983), however, a rigorous verification of (a) and (b) would appear to be far from easy in general, although one would expect such results to be true for many cases of interest.

It should be noted that the approach considered here is not the same as that considered by Keiding (1974) for the birth process, and later more generally by Basawa and Brockwell (1984). They condition on the *limit* random variable W,

Received April 1985; revised July 1985.
AMS 1980 *subject classifications*. Primary 60J80, 62F12.
Key words and phrases. Asymptotic conditional inference, nonergodic models, supercritical branching process, maximum likelihood estimator, asymptotic ancillarity.

and then treat the unobserved value w of W as a nuisance parameter, which is then estimated, via V_n for instance. This approach has the attraction of reducing a nonergodic model to an ergodic one; the final result still does not tell us whether asymptotic normality holds *conditional* on V_n, however.

In this paper we obtain the conditional limit theorem in the case of a supercritical Galton–Watson process with unknown offspring mean θ, when the offspring distribution is a member of the power series family of distributions. The "related" conditioning statistic here is the total number of ancestors V_n, which is proportional to the conditional information. Moreover, the results imply that V_n is asymptotically ancillary for θ in a local sense. The derivation, which follows the detailed analysis in Dubuc and Seneta (1976), highlights the difficulties involved in establishing such conditional limit theorems more generally. The problem is rather more difficult than, and in fact includes, the problem of establishing a local limit result for V_n.

There have been a number of articles recently concerning approximate ancillarity and conditionality, mainly pertaining to independent samples. Much of this work has developed from the papers by Efron and Hinkley (1978) and Cox (1980); see also Hinkley (1980), Barndorff-Nielsen (1980), Ryall (1981), Amari (1982), and McCullagh (1984). In particular, the construction of approximate ancillary statistics based on the observed information has received much attention.

2. Preliminaries and statement of results. Let $(Z_0 = 1, Z_1, \ldots, Z_n)$ be a sample of successive generation sizes from a supercritical Galton–Watson process. We assume that the nondegenerate offspring distribution (p_j) is a member of the power series family, $p_j = f_j \lambda^j \{F(\lambda)\}^{-1}$ where $\lambda > 0$, $f_j \geq 0$, and $F(\lambda) = \sum_j f_j \lambda^i$. We suppose that $f_0 = 0$, so that $\theta = E(Z_1) > 1$, and assume $\sigma^2 \equiv \sigma^2(\theta) = \mathrm{Var}(Z_1) < \infty$. The maximum likelihood estimator $\hat{\theta}_n$ of θ is given by

$$\hat{\theta}_n = 1 + V_n^{-1}(Z_n - 1),$$

where $V_n = \sum_{j=0}^{n-1} Z_j$ is the total number of ancestors (Heyde, 1975). Moreover, the conditional information is $\sigma^{-2} V_n$ (Heyde, 1975), and

$$T_n(\theta) \equiv (\theta^n - 1)^{-1}(\theta - 1) V_n \to W \quad \text{a.s.}$$

It is shown by Basawa and Scott (1976) that

$$(X_n(\theta), T_n(\theta)) \to (Z, W) \tag{1}$$

in joint distribution, where $X_n(\theta) = V_n^{1/2}(\hat{\theta}_n - \theta)$ and $Z \sim N(0, \sigma^2)$ independently of W. Furthermore, the convergence in (1) is locally uniform in $\theta > 1$ [Sweeting (1978, 1980)].

As discussed in the previous section, if V_n could be regarded as "asymptotically ancillary" for θ, then the conditionality principle would dictate basing inferences about θ on the conditional distribution of $\hat{\theta}_n$ given V_n. Moreover, V_n is a prime candidate for such a conditionality resolution as it directly affects the *precision* of $\hat{\theta}_n$; see for example Efron and Hinkley (1978) and the related discussion. A statistic V_n is often said to be approximately ancillary if its distribution is approximately free from θ. In practice, this usually comes down to checking that the density of V_n is approximately free from θ, via an Edgeworth expansion, for example, in the independent case. The *definition* of asymptotic ancillarity however should really be based on the asymptotic behaviour of the density or, as in this case, the probability mass function (p.m.f.) of V_n, since it is the lack of information contained in the observed value of V_n which is relevant. One possible formulation of asymptotic ancillarity in a local sense of V_n here is

$$\lim_{n\to\infty} \frac{P(V_n = v_n|\theta_n)}{P(V_n = v_n|\theta)} = 1, \tag{2}$$

whenever $|\theta_n - \theta| \le A\theta^{-1/2n}$ and $v_n/c_n \to w$ for any $A > 0$, $w > 0$, where $c_n \equiv c_n(\theta) = (\theta - 1)^{-1}(\theta^n - 1)$. [It should be noted that in Cox (1980) the phrase "local ancillarity" refers to the behaviour of the *distribution* in a neighbourhood of the true value of θ.] A very similar criterion to (2) was also used in Section 5, Chapter 4 of Basawa and Scott (1983) while investigating the efficiency of conditional tests in mixed exponential families.

Borrowing the notation in Dubuc and Seneta (1976), we shall say that the process is of type (L, r) if L is the greatest integer for which the offspring distribution is defined on a lattice $\{kL + r: k = 0, 1, \dots\}$. We prove the following result.

THEOREM. (i) *V_n is asymptotically ancillary for θ in the sense of* (2), *and* (ii) *if (v_n) is a sequence of integers such that $v_n = \sum_{j=0}^{n-1} r^j (\mathrm{mod}\ L)$ and $v_n/c_n \to w > 0$ then*

$$v_n^{1/2}(\hat{\theta}_n - \theta)|V_n = v_n \to Z$$

in distribution, where $Z \sim N(0, \sigma^2)$, uniformly in compact intervals of $\theta \in (1, \infty)$.

The proof is via a detailed analysis of the joint characteristic function of (Z_n, V_n), along the lines of Dubuc and Seneta (1976). As a by-product, the local limit theorem for V_n is established, which implies the asymptotic ancillarity of V_n.

Write $S_n = (Z_n - 1) - (\theta - 1)V_n$ and let $U_n = c_n^{-1/2}S_n = T_n^{1/2}X_n$ (suppressing the parameter θ). Define the characteristic functions $\phi_n(\chi, \eta) = E(e^{i(\chi S_n + \eta V_n)})$, $\psi_n(\zeta, \xi) = E(e^{i(\zeta U_n + \xi T_n)}) = \phi_n(c_n^{-1/2}\zeta, c_n^{-1}\xi)$. From (1) and the continuous mapping theorem we have

$$\psi_n(\zeta, \xi) \to \psi(\zeta, \xi) = E(e^{i(\zeta U + \xi W)}) \tag{3}$$

uniformly in finite rectangles, where $U = W^{1/2}Z$. Furthermore, $\psi(\zeta, \xi) = Ee^{i\xi W}E(e^{i\zeta W^{1/2}Z}|W) = g((\sigma\zeta)^2 - i\xi)$ where $g(s) = E(e^{-sW})$. Let $\psi_n^v(\zeta) = E(e^{i\zeta U_n}|V_n = v)$ be the conditional characteristic function of U_n given $V_n = v$. The relationship between $\psi_n^v(\zeta)$ and the joint characteristic function is given by

$$\psi_n^v(\zeta)p_n^0(v) = p_n^\zeta(v),$$

where

$$p_n^\zeta(v) = (L/2\pi)\int_{-\pi/L}^{\pi/L} e^{-iv\xi}\psi_n(\zeta, c_n\xi)\, d\xi.$$

See for example Bartlett (1938). Clearly $p_n^0(v)$ is the p.m.f. of V_n. The proof consists essentially of showing that $L^{-1}c_n p_n^\zeta(v_n) \to p^\zeta(w)$ for an appropriate sequence (v_n) with $c_n^{-1}v_n \to w > 0$, where $p^\zeta(w) = (1/2\pi)\int_{-\infty}^{\infty} e^{-\iota w\xi}\psi(\zeta, \xi)\, d\xi$, from which it will follow that $\psi_n^v(\zeta) \to \psi^w(\zeta) \equiv E(e^{\iota U\zeta}|W = w)$.

3. Lemmas. We need a number of results concerning the joint characteristic function $\phi_n(\chi, \eta)$ of (S_n, V_n). Let $H_n(s, t) = E(s^{Z_n}t^{V_n})$, $n \geq 1$, be the joint probability generating function of (Z_n, V_n). These generating functions are recursively related by

$$H_n(s, t) = tf(H_{n-1}(s, t)), \qquad n \geq 1, \quad H_0(s, t) = s, \tag{4}$$

where $f(s) = E(e^{sZ_1})$ [Jagers (1975)]. Define $K(z, \theta) = e^{i\theta}f(z)$ where θ is real and z complex with $|z| \leq 1$. Let $K_n(z, \theta)$ be the nth functional iterate of $K(z, \theta)$ in the first argument: that is, $K_n(z, \theta) = K(K_{n-1}(z, \theta), \theta)$, $n \geq 1$, where we have set $K_0(z, \theta) = z$. It follows that $H_n(s, e^{\iota\eta}) = K_n(s, \eta)$, and so $\phi_n(\chi, \eta) = H_n(e^{i\chi}, e^{i(\eta-(\theta-1)\chi)}) = e^{-i\chi}K_n(e^{\iota\chi}, \eta - (\theta - 1)\chi)$. We shall need the following estimates.

Lemma 1. *There exists* ρ *with* $p_1 < \rho < 1$ *such that for all* $p \geq 1$, $|z|$, $|z'| \leq R < 1$ *and all* η, η' *one has*

$$\left|K_p(z, \eta)\right| \leq A_R\rho^p \tag{5}$$

and

$$\left|K_p(z, \eta) - K_p(z', \eta')\right| \leq B_R\rho^p(|z - z'| + |\eta - \eta'|). \tag{6}$$

Proof. Choose h sufficiently small for $p \equiv f'(h) < 1$. If $0 < R < 1$ then $f_p(R) \downarrow 0$ where $f_p(s) = H_p(s, 1)$, and so there exists an integer N such that $f_p(R) \leq h$ for all $p \geq N$. Then if $|z| \leq R < 1$ and $p > N$,

$$\left|K_p(z, \eta)\right| = \left|K\left(K_{p-1}(z, \eta), \eta\right)\right| \leq \left|f'(z_1)\right| \left|K_{p-1}(z, \eta)\right|,$$

where $|z_1| \leq |K_{p-1}(z, \eta)| \leq h$, since $K(0, \eta) = 0$. Thus $|K_p(z, \eta)| \leq \rho^{p-N}|K_N(z, \eta)|$ and (5) follows.

For (6), write $\delta_p = |K_p(z, \eta) - K_p(z', \eta')|$; then if $p > N$

$$\begin{aligned}\delta_p &\leq \left|f\left(K_{p-1}(z, \eta)\right) - f\left(K_{p-1}(z', \eta')\right)\right| + \left|e^{\iota(\eta-\eta')} - 1\right| \left|K_p(z', \eta')\right| \\ &\leq \rho\delta_{p-1} + |\eta - \eta'|A_R\rho^p\end{aligned}$$

from (5). Iterating, one arrives at $\delta_p \leq \rho^{p-N}(\delta_N + A_R(1 - \rho)^{-1}|\eta - \eta'|)$. Finally, it is readily verified that $|(\partial/\partial z)K_N(z, \eta)| < C_1$ and $|(\partial/\partial\eta)K_N(z, \eta)| < C_3$ for all $|z| < 1$ and all η, so that $\delta_N \leq C_2|z - z'| + C_3|\eta - \eta'|$ and (6) follows. □

Lemmas 2–7 mirror results in Section 2 of Dubuc and Seneta (1976) for the characteristic function of Z_n.

Lemma 2. *For all* $\varepsilon > 0$

$$\sup\left\{\left|\phi_n(\chi, \eta)\right|: n \geq 1, |\chi| \leq \pi/L, \varepsilon/c_n \leq |\eta| \leq \pi/L\right\} < 1.$$

PROOF. From (3), $\psi_n(\zeta, \xi) \to g((\sigma\zeta)^2 - i\xi)$ uniformly in finite intervals of $\mathbb{R}^2$. Clearly $|g((\sigma\zeta)^2 - i\xi)| \le g((\sigma\zeta)^2)$, and $|g(-i\xi)| < 1$ if $\xi \ne 0$, as in the proof of Lemma 1 in Dubuc and Seneta (1976). Since $g(s)$ is a decreasing function of $s > 0$ it follows that $\sup\{|\psi(\zeta, \xi)|: \zeta \in R\} < 1$ for each $\xi \ne 0$. Thus there exists $r \in (0,1)$ and $N > 1$ such that $|\psi_n(\zeta, \xi)| \le r$ if $n \ge N$ and $\varepsilon \le |\xi| \le (1+\theta)\varepsilon$. Then if $\varepsilon/c_k \le |\eta| \le \varepsilon/c_{k-1}$ we have $|\phi_k(\chi, \eta)| \le r$ for $k \ge N$ since $c_k/c_{k-1} \le 1 + \theta$, and hence for this range of values

$$\begin{aligned}|\phi_n(\chi,\eta)| &= |K_n(e^{i\chi}, \eta - (\theta-1)\chi)| \le |K_k(e^{i\chi}, \eta - (\theta - 1)\chi)| \\ &= |\phi_k(\chi,\eta)| \le r.\end{aligned}$$

Consider finally the region $\varepsilon/c_{n-1} \le |\eta| \le \pi/L$. When $0 < \delta \le |\gamma| \le \pi/L$ there exists $S(\delta) < 1$ such that $|f(e^{i\gamma})| \le S(\delta)$. Thus $|\phi_n(\chi, \eta)| \le |\phi_1(\chi, \eta)| = |f(e^{i\chi})| \le S(\delta) < 1$ provided $0 < \delta < |\chi| \le \pi/L$. When $\chi = 0$ we have $|\phi_n(0, \eta)| \le |\phi_2(0, \eta)| = |f(e^{i\eta})| \le S(\varepsilon/c_{N-1}) < 1$ and it follows by continuity that $\sup\{|\phi_n(\chi, \eta)|: |\chi| \le \pi/L,\ \varepsilon/c_{N-1} \le |\eta| \le \pi/L\} < 1$ as required. □

In a similar way to Dubuc and Seneta (1976) we define the sequence of intervals

$$J_k = \{\eta: \pi L^{-1}c_k^{-1} \le |\eta| \le \pi L^{-1}c_{k-1}^{-1}\}, \qquad k > 1.$$

LEMMA 3. *For all $n \ge k$, $|\chi| \le \pi/L$, $\eta \in J_k$, $k > 1$ there exists a constant A such that*

$$|\phi_n(\chi, \eta)| \le A\rho^{n-k}.$$

PROOF. We have

$$\begin{aligned}|\phi_n(\chi,\eta)| &= |K_n(e^{i\chi}, \eta - (\theta-1)\chi)| \\ &= |K_{n-k}(e^{i\chi}\phi_k(\chi,\eta), \eta - (\theta-1)\chi)|\end{aligned}$$

and by Lemma 2 there exists $R < 1$ such that $|\phi_k(\chi, \eta)| \le R$ for all $|\chi| \le \pi/L$, $\eta \in J_k$, $k > 1$. The result now follows immediately from inequality (5) of Lemma 1. □

Dubuc and Seneta (1976) show that $g(s)$ is a lower bound for $f_n(e^{-s/c_n})$ for all $s \in (0,1)$, where $f_n(s) = H_n(s, 1)$. We require a similar bound for $h_n(e^{-t/c_n})$, where $h_n(t) = H_n(1, t)$. Establishing such a bound requires a little more work, but one does arise as a consequence of the bound for $f_n(e^{-s/c_n})$.

LEMMA 4. *There exist numbers $\alpha > 0$, $\tau < 1$ such that for all $t \in (0, \tau)$ and all $n \ge 1$, $h_n(e^{-t/c_n}) \ge g(\alpha t)$.*

PROOF. Write $P(s) = -\log f(e^{-s})$. Then $P(s)$ is an increasing concave function, $P(0) = 0$, $\lim_{s\to\infty} P(s) = \infty$, and $P_n(s) \equiv -\log f_n(e^{-s})$ is the nth functional iterate of $P(s)$. Write $Q_n(t) = -\log h_n(e^{-t})$; then from (4), $Q_n(t) =$

$P(Q_{n-1}(t)) + t$, $n \geq 1$, from which it follows that $Q_{n-1}(t) + t \leq Q_n(t)$, since $P'(s) > 1$ for all $s > 0$. Let $s_0 > 0$ and let $\lambda = P'(s_0) > 1$. Then if $u + t/\lambda \leq s_0$ a first-order Taylor expansion gives $P(u) + t \leq P(u + t/\lambda)$. Now choose t so small that $Q_n(t) \leq s_0$. It then follows that $Q_n(t) = P(Q_{n-1}(t)) + t \leq P(Q_{n-1}(t) + t/\lambda)$ since $Q_{n-1}(t) + t/\lambda \leq Q_{n-1}(t) + t \leq Q_n(t) \leq s_0$. Iterating, one finds that

$$Q_n(t) \leq P_n\big(Q_0(t) + t(\lambda^{-1} + \cdots + \lambda^{-(n-1)})\big) \leq P_n(\alpha t),$$

where $\alpha = \lambda/(\lambda - 1)$. Finally, since $Q_n(t/c_n) \to -\log g(t)$ one can choose t_1 so small that $Q_n(t/c_n) \leq s_0$ for all $n \geq 1$ and $0 < t \leq t_1$, giving $Q_n(t/c_n) \leq P_n(\alpha t/c_n)$. Therefore $h_n(e^{-t/c_n}) \geq f_n(e^{-\alpha t/c_n}) \geq g(\alpha t)$ if $0 < t < \alpha^{-1}$, from Lemma 5 in Dubuc and Seneta (1976), and the result follows on choosing $\tau = \min(\alpha^{-1}, t_1)$. □

LEMMA 5. *There exists a function V^* defined on $(0, \infty)$ such that* (a) *V^* is slowly varying as $x \to 0+$*, (b) *V^* is bounded on every interval (ε, ∞), $\varepsilon > 0$, and* (c) *for all $|\chi| < \pi/L$, $\eta, \eta' \in J_k$ with $|\eta - \eta'| \leq \delta$ and all $n \geq k$*

$$|\phi_n(\chi, \eta) - \phi_n(\chi, \eta')| \leq c_k \delta \rho^{n-k} V^*(c_k\delta).$$

PROOF. We have

$$\begin{aligned}|\phi_n(\chi, \eta) - \phi_n(\chi, \eta')| &= |K_n(e^{i\chi}, \eta - (\theta - 1)\chi) - K_n(e^{i\chi}, \eta' - (\theta - 1)\chi)| \\ &= |K_{n-k}(z, \beta) - K_{n-k}(z', \beta')|,\end{aligned}$$

where $z = e^{i\chi}\phi_k(\chi, \eta)$, $\beta = \eta - (\theta - 1)\chi$, etc. From Lemma 2 there exists $R < 1$ such that $|\phi_k(\chi, \eta)| \leq R$ for all $n \geq k$, $\eta \in J_k$, and Lemma 1 now gives

$$(7) \qquad |\phi_n(\chi, \eta) - \phi_n(\chi, \eta')| \leq B_R \rho^{n-k}\big(|\phi_k(\chi, \eta) - \phi_k(\chi, \eta')| + \delta\big).$$

But

$$\begin{aligned}|\phi_k(\chi, \eta) - \phi_k(\chi, \eta')| &= \big|E\big(e^{i(\chi Z_k + \eta' V_k)}(e^{i(\eta - \eta')V_k} - 1)\big)\big| \\ &\leq E|e^{i(\eta - \eta')V_k} - 1| \leq C\big(1 - h_k(e^{-\delta})\big)\end{aligned}$$

as in the proof of Lemma 6 in Dubuc and Seneta (1976). But if $c_k\delta \leq \tau$ then Lemma 4 of that paper and Lemma 4 here give

$$1 - h_k(e^{-\delta}) \leq \alpha c_k \delta V(c_k\delta),$$

where $V(s) = (1 - g(s))/s$ is slowly varying as $s \to 0+$. Thus, $|\phi_k(\chi, \eta) - \phi_k(\chi, \eta')| \leq c_k\delta V_1(c_k\delta)$ where $V_1(x) = C\alpha V(x)$, $x \leq \tau$ and $V_1(x) = C$, $x > \tau$. With a suitable choice of $V^*(x)$ the result follows from (7). □

LEMMA 6. *Let $\phi(\zeta, \xi)$ be the joint characteristic function of (X, Y), and suppose that*

$$(1/2\pi)\int_{-\tau_n}^{\tau_n} \phi(\zeta, \xi) e^{-iy\xi}\, d\xi$$

converges locally uniformly in y to a function $p^{\zeta}(y)$ (necessarily continuous in y)

for every fixed ζ and any fixed sequence of positive numbers $\tau_n \to \infty$ as $n \to \infty$. Then $p^0(y)$ is the density of Y, and $p^\zeta(y) = \phi^y(\zeta)p^0(y)$ where $\phi^y(\zeta) = E(e^{i\zeta X}|Y = y)$.

PROOF. Consider the complex measure $dF^\zeta(y) = \phi^y(\zeta)\,dF(y)$, where F is the distribution of Y. Note that F^ζ is of bounded variation, and that the characteristic function ψ^ζ of F^ζ is given by $\psi^\zeta(\xi) = \int e^{i\xi y}\phi^y(\zeta)\,dF(y) = \phi(\zeta, \xi)$. Then, as in Dubuc and Seneta (1976), it is seen that $\int_a^b p^\zeta(y)\,dy = F^\zeta(a, b)$ at continuity points of F^ζ, the proof applying without change to a complex measure. Thus $p^\zeta(y)$ is the density of F^ζ. Setting $\zeta = 0$, it follows that $p^0(y)$ is the density of Y, and hence $p^\zeta(y) = \phi^y(\zeta)p^0(y)$ as required. □

LEMMA 7.

$$(1/2\pi)\int_{-\pi c_n/L}^{\pi c_n/L} e^{-iw\xi}\psi_n(\zeta, \xi)\,d\xi \to \psi^w(\zeta)p(w)$$

locally uniformly in $\zeta \in R$ and $w > 0$ where $p(w)$ is the continuous density of W, and $\psi^w(\zeta) = E(e^{i\zeta U}|W = w)$.

PROOF. The argument here is identical to that given in Dubuc and Seneta (1976) on taking $K_n(\xi, w) = e^{-iw\xi}\psi_n(\zeta, \xi)$, $q = 0$ and using Lemmas 3, 5, and 6 here in place of their Lemmas 3, 7, and 8, and we omit the details. It is only necessary to note that $c_n^{-1/2}|\zeta| < \pi/L$ for all n sufficiently large for the applicability of Lemmas 3 and 5. □

4. Proof of the conditional limit theorem. If $Z = X + Y$ where X, Y are two lattice random variables such that the distributions of X and $Y|X = x$ have period L, then it is easily seen that the distribution of Z must also have period L. It follows that the distribution of both Z_n and V_n have period L for all $n \geq 1$. The possible values of V_n are easily seen to be among $\sum_{i=0}^{n-1} r^i (\mathrm{mod}\ L)$ and hence if (v_n) is any sequence of positive integers such that $v_n = \sum_{i=0}^{n-1} r^i (\mathrm{mod}\ L)$ and $\lim_{n\to\infty} v_n/c_n = w > 0$ then we have [cf. Bartlett (1938) and Steck (1957)] $\psi_n^{v_n}(\zeta)p_n^0(v_n) = p_n^\zeta(v_n)$ and

$$L^{-1}c_n p_n^\zeta(v_n) = (1/2\pi)\int_{-\pi c_n/L}^{\pi c_n/L} e^{-i(v_n/c_n)\xi}\psi_n(\zeta, \xi)\,d\xi.$$

But now from Lemma 7, $L^{-1}c_n p_n^\zeta(v_n) \to \psi^w(\zeta)p(w)$ locally uniformly in $\zeta \in \mathbb{R}$. We have therefore shown that $\psi_n^{v_n}(\zeta) \to \psi^w(\zeta) = E(e^{i\zeta U}|W = w)$ locally uniformly in $\zeta \in \mathbb{R}$.

Finally,

$$E\left(e^{i\zeta X_n}|V_n = v_n\right) = E\left(e^{i\zeta(v_n/c_n)^{-1/2}U_n}|V_n = v_n\right)$$
$$\to E\left(e^{i\zeta w^{-1/2}U}|W = w\right) = E(e^{i\zeta Z})$$

and the convergence in (ii) follows for each $\theta > 1$. For (i) and the uniformity in

(ii) we need to check that the convergence in Lemma 7 is uniform in compact intervals of $(1, \infty)$. The convergence $\psi_n(\zeta, \xi) \to \psi(\zeta, \xi)$ required in Lemmas 2 and 7 is uniform in compacts of $(1, \infty)$ from the uniform convergence in (1). Finally, the choice of constants ρ, A_R, and B_R in the bounds (5) and (6) may be made independently of θ in compact intervals of $(1, \infty)$ for reasons of continuity.

Taking $\zeta = 0$ we see that

$$\frac{P(V_n = v_n|\theta_n)}{P(V_n = v_n|\theta)} = \frac{L^{-1}c_n(\theta_n)p_n^0(v_n|\theta_n)}{L^{-1}c_n(\theta)p_n^0(v_n|\theta)}\frac{c_n(\theta)}{c_n(\theta_n)} \to 1$$

provided $c_n(\theta_n)/c_n(\theta) \to 1$ as $n \to \infty$, which will be the case if and only if $\theta_n - \theta = o(n^{-1})$, and (2) follows. □

5. Concluding remarks. For ease of exposition, only the case $X_0 = 1$ was treated here. It is a relatively straightforward matter to show that the theorem remains true in the general case when $X_0 = j \geq 1$. [In the case of an arbitrary initial distribution $(a_j,\ j \geq 1)$, then usually one would want to condition on the observed value X_0, provided of course that (a_j) is independent of θ.]

A more substantial generalization would be to relax the assumption that $p_0 = 0$. In this case, it will be necessary to argue conditionally on nonextinction of the process by time n. (It is inappropriate in the author's view to condition on nonextinction of the entire process, as this information is never actually available.) This would require an extension of the arguments given here to the case where $p_0 > 0$, and this has not been attempted. Note however that since $P(X_n > 0|\theta) \to 1 - q(\theta)$ locally uniformly in θ where $q(\theta)$ is the probability of ultimate extinction, and $q(\theta)$ is continuous, the event $\{X_n > 0\}$ is asymptotically ancillary in the sense used here.

A referee has pointed out that there must be a connection between the asymptotic ancillarity of V_n as defined here and a concept of asymptotic ancillarity defined in terms of Fisher's information contained in V_n. Specifically, let $I_n(\theta)$ be Fisher's information in the observed process up to time n and $I_{V_n}(\theta) = E[k_n'(\theta)]^2$ where $k_n(\theta) = \log P(V_n = v_n|\theta)$. Then V_n is asymptotically ancillary in this sense if $I_{V_n}(\theta)/I_n(\theta) \to 0$ as $n \to \infty$. Indeed, in the present problem this ratio is of order $n^2\theta^{-n}$. In the case of independence, Amari (1982) defines higher-order asymptotic ancillarity essentially in terms of the order of the corresponding ratio of information functions. It can be seen informally that the two approaches to asymptotic ancillarity are very close, as $k_n(\theta + \delta I_n^{-1/2}) - k_n(\theta) \approx \delta I_n^{-1/2}k_n'(\theta)$. Thus the convergence of I_{V_n}/I_n to zero will usually entail (2) and vice versa. For a formal result it will be necessary to impose further conditions on the sequence $I_n^{-1/2}k_n'(\theta)$, and the question is not pursued further here. Nevertheless, the close connection between the two concepts is illuminating. The rate at which the ratio (2) converges to one is a measure of the degree of ancillarity; in our case it is $O(n\theta^{-1/2n})$, which is the same order as $(I_{V_n}/I_n)^{1/2}$.

Acknowledgment. The author wishes to express his thanks to the referees for their constructive comments—in particular those relating to the concept of asymptotic ancillarity.

REFERENCES

AMARI, S. (1982). Geometrical theory of asymptotic ancillarity and conditional inference. *Biometrika* **69** 1–17.

BARNDORFF-NIELSEN, O. (1980). Conditionality resolutions. *Biometrika* **67** 293–310.

BARTLETT, M. S. (1938). The characteristic function of a conditional statistic. *J. London Math. Soc.* **13** 62–67.

BASAWA, I. V. and BROCKWELL, P. J. (1984). Asymptotic conditional inference for regular nonergodic models with an application to autoregressive processes. *Ann. Statist.* **12** 161–171.

BASAWA, I. V. and SCOTT, D. J. (1976). Efficient tests for branching processes. *Biometrika* **63** 531–536.

BASAWA, I. V. and SCOTT, D. J. (1983). *Asymptotic Optimal Inference for Nonergodic Models. Lecture Notes in Statist.* **17**. Springer, Berlin.

COX, D. R. (1980). Local ancillarity. *Biometrika* **67** 279–286.

DUBUC, S. and SENETA, E. (1976). The local limit theorem for the Galton–Watson process. *Ann. Probab.* **4** 490–496.

EFRON, B. and HINKLEY, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information (with discussion). *Biometrika* **65** 457–487.

FEIGIN, P. D. and REISER, B. (1979). On asymptotic ancillarity and inference for Yule and regular nonergodic processes. *Biometrika* **66** 279–283.

HEYDE, C. C. (1975). Remarks on efficiency in estimation for branching processes. *Biometrika* **62** 49–55.

HINKLEY, D. V. (1980). Likelihood as approximate pivotal distribution. *Biometrika* **67** 287–292.

JAGERS, P. (1975). *Branching Processes with Biological Applications.* Wiley, New York.

KEIDING, N. (1974). Estimation in the birth process. *Biometrika* **61** 71–80.

MCCULLAGH, P. (1984). Local sufficiency. *Biometrika* **71** 233–244.

RYALL, T. A. (1981). Extensions of the concept of local ancillarity. *Biometrika* **68** 677–683.

STECK, G. P. (1957). Limit theorems for conditional distributions. *Univ. Calif. Publ. Stat.* **42** 237–284.

SWEETING, T. J. (1978). On efficient tests for branching processes. *Biometrika* **65** 123–127.

SWEETING, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8** 1375–1381.

SWEETING, T. J. (1982). Correction note to "Uniform asymptotic normality of the maximum likelihood estimator." *Ann. Statist.* **10** 320–321.

SWEETING, T. J. (1983). On estimator efficiency in stochastic processes. *Stochastic Process. Appl.* **15** 93–98.

DEPARTMENT OF STATISTICS
UNIVERSITY OF SURREY
GUILDFORD SURREY GU2 5XH
ENGLAND